

**CLAIMS:**

1. Search apparatus for searching for data in the form of units of a natural language, the apparatus comprising:
  - 5 interface means for receiving an input query in the form of units of the natural language and for outputting the results of the search in the form of output data;
  - 10 matching means for searching for and identifying any matches between the units of the input query and the units of the data;
  - 15 generating means for, where there are unmatched units in the query and/or the data, generating context data in the form of one or more unmatched units of the query and/or the data, each unmatched unit having a predefined linguistic relationship to one of the or each matched unit; and
  - 20 forming means for forming said output data as any said matched units and any respective said context data.
- 25 2. Search apparatus according to claim 1, wherein said generating means is adapted to generate the or each unmatched unit of the context data having a predefined modification relationship to the respective matched units.
3. Search apparatus according to claim 1, wherein said generating means operates in accordance with one or more

rules defining contextually important modification relationships between matched and unmatched units.

4. Search apparatus according to claim 1, wherein said generating means operates in accordance with one or more rules containing syntactic and semantic constraints for the formation of the context data.
5. Search apparatus according to claim 1, including parsing means for parsing the input query and/or the data or the output of the matching means to determine linguistic relationships between the units.
6. Search apparatus according to claim 1, wherein said forming means is adapted to form said output data as any said matched units associated with respective said unmatched units having said predefined linguistic relationship to respective matched units.
- 20 7. Search apparatus according to claim 1, wherein said forming means is adapted to form said output data as a layered hierarchical structure identifying sets of data by their context data.
- 25 8. Search apparatus according to claim 7, wherein said forming means is adapted to form said output data as a hierarchical structure formed from a said matched word comprising a head word of the input query, said context

data for said head word forming one or more sublayers, any further matched words forming further sublayers of said sublayers, the order of selection of said further matched words being dependent on their modification relationship within the input query, said context data for said further matched words forming yet further sublayers, said sets of data being identified by a final said sublayer in the hierarchical structure.

10 9. A computer implemented data processing method for processing data to enhance the results of a search for data in the form of units of a natural language, the method comprising:

15 receiving an input query in the form of units of the natural language and outputting the results of the searching the form of output data;

searching for and identifying any matches between the units of the input query and units of the data;

20 for any matched units in the query and/or the data, generating context data in the form of one or more unmatched units of the query and/or the data, each unmatched unit having a predefined linguistic relationship to one of the or each matched unit; and

25 forming said output data as any said matched units and any respective said context data.

10. A method according to claim 9, wherein the or each unmatched unit of the context data is generated having a

predefined modification relationship to the respective matched units.

- 5        11. A method according to claim 9, wherein the context data is generated in accordance with one or more rules defining contextually important modification relationships between matched and unmatched units.
- 10      12. A method according to claim 9, wherein the context data is generated in accordance with one or more rules containing syntactic and semantic constraints for the formation of the context data.
- 15      13. A method according to claim 9, including parsing the input query and/or the data or the output of the matching step to determine linguistic relationships between the units.
- 20      14. A method according to claim 9, wherein the output data is formed as any said matched units associated with respective said unmatched units having said predefined linguistic relationships to respective matched units.
- 25      15. A method according to claim 9, wherein said output data is formed as a layered hierarchical structure identifying sets of data by their context data.

16. A method according to claim 15, wherein said output data is formed from a said matched word comprising a head word of the input query, said context data for said head word forming one or more sublayers, any further matched words forming further sublayers of said sublayers, the order of selection of said further matched words being dependent on their modification relationship within the input query, said context data for said further matched words forming yet further sublayers, said sets of data being identified by a final said sublayer in the hierarchical structure.

17. Data retrieval apparatus for retrieving desired information units of a natural language from a plurality of available information units, the apparatus comprising:

input means for inputting a query in units of the natural language;

matching means for searching for and identifying any matches between the units of the input query and the units of the available information units to identify the best matches between the input query and the plurality of available information units;

generating means for receiving the best matches and for where there are unmatched units in the input query and/or the information units, generating context data in the form of one or more unmatched units each having a predefined linguistic relationship to one of the or each matched unit; and

output means for outputting desired information units as the best matches with respective context data.

18. Data retrieval apparatus according to claim 17,  
5 wherein said output means is adapted to output the desired information units ordered by said context data.

19. A computer implemented data retrieval method for retrieving desired information units containing units of  
10 a natural language for a plurality of available information units, the method comprising:

    inputting a query in units of the natural language;  
    searching for and identifying any matches between  
    the units of the input query and the units of the  
15 available information units to identify the best matches between the input query and the plurality of available information units;

    for the best matches where there are unmatched units in the query and/or the information units, generating  
20 context data in the form of one or more unmatched unit each having a predefined linguistic relationship to one of the or each matched unit; and

    outputting desired information units as the best matches with respective said context data.

25

20. The method of claim 19, wherein the desired information units are output ordered by said context data.

21. A carrier medium carrying processor implementable instructions for controlling a processor to carry out the method of any one of claims 9 to 16, 19 or 20.
- 5 22. A signal carrying processor implementable instructions for controlling a processor to carry out the method of any one of claims 9 to 16, 19 or 20.